(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
27 October 2005 (27.10.2005)

PCT

(10) International Publication Number
WO 2005/101292 A2

(51) International Patent Classification[7]: G06K 9/00

(21) International Application Number: PCT/FR2005/000673

(22) International Filing Date: 18 March 2005 (18 03 2005)

(25) Filing Language: French

(26) Publication Language: French

(30) Priority Data:
04/03,556 5 April 2004 (05 04 2004) FR

(71) Applicant and
(72) Inventor : LEBRAT, François [FR/FR]; 98 avenue de Versailles, F-75016 Paris (FR)

(74) Agents: HASSINE, Albert; Cabinet Plasseraud, 65/67 rue de la Victoire, F-75440 Paris Cedex 09 (FR)

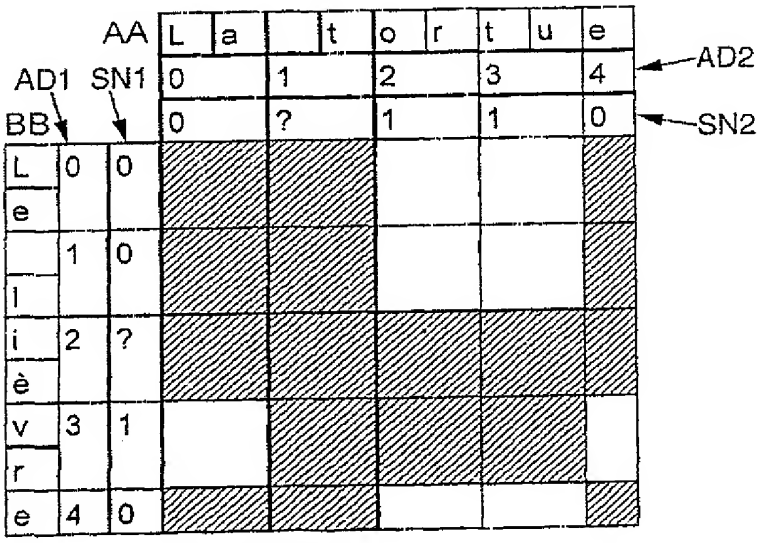
(81) Designated states *(unless otherwise indicated for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW

[continued on next page]

As printed

(54) Title: METHOD FOR SEARCHING CONTENT PARTICULARLY FOR EXTRACTS COMMON TO TWO COMPUTER FILES

(54) Titre : PROCÉDÉ DE RECHERCHE DE CONTENU NOTAMMENT D'EXTRAITS COMMUNS ENTRE DEUX FICHIERS INFORMATIQUES



(57) Abstract: The present invention relates to searching content particularly for at least one extract common to a first data file and a second data file. The method comprises a preliminary step of preparing at least the first file by (a) dividing the first file into a series of data packets having a predetermined size and identifying packet addresses in said file (b) combining each packet address with a digital signature that defines one of three fuzzy logic states namely true, false and indeterminate, and is the result of a combinatorial computation on data from said file; whereafter said method comprises performing an actual search for a common extract by (c) comparing the fuzzy logic states combined with each packet address of the first file with fuzzy logic states determined on the basis of data from the second file, and (d) removing from said common extract search the respective address pairs from the first and second files that have the respective logic states true and false or false and true and retaining the other address pairs that identify data packets that may comprise said common extract

WO 2005/101292 A2

[continued on next page]

(84) **Designated states** *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)

Declaration under Rule 4 17:

- *of inventorship (Rule 4 17.iv)) for the following designation US*

Published:

- *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette

(57) Abrégé : L'invention concerne la recherche de contenu, notamment une recherche d'au moins un extrait commun entre un premier fichier et un second fichier de données L e procédé de l'invention comporte une préparation préalable du premier fichier au moins comprenant les étapes suivantes : a) segmenter le premier fichier en une succession de paquets de données de taille choisie, et identifier des adresses de paquets dans ledit fichier b) associer à l'adresse de chaque paquet une signature numérique définissant un état en logique floue parmi au moins trois états : "vrai", faux" et "indéterminé", ladite signature résultant d'un calcul combinatoire sur des données issues dudit fichier, et le procédé se poursuit par une recherche d'extrait commun, proprement dite, comprenant les étapes suivantes, c) comparer les états de logique floue associés à chaque adresse de paquet du premier fichier avec des états de logique floue déterminés à partir de données issues du second fichier d) éliminer de ladite recherche d extrait commun des couples d adresses respectives des premier et second fichiers dont les états logiques respectifs sont 'vrai et faux ou 'faux' et 'vrai", et conserver les autres couples d'adresses identifiant des paquets de données susceptibles de comporter ledit extrait commun